Customer Behaviour Analysis

1. Project Overview

This project focuses on analyzing customer purchasing behaviour to understand buying patterns and preferences. The dataset was cleaned and transformed using Python to ensure accuracy and consistency. Power BI was used to create interactive dashboards for visual analysis. The analysis helps identify trends related to age group, gender, product categories, and spending behaviour. These insights can support better business decision-making and customer targeting strategies.

Exploratory Data Analysis using Python

We began with data preparation and cleaning in Python:

- Data Loading: Imported the dataset using pandas.

```
df = pd.read_csv("C:/Users/Simran/Downloads/jupyter_notebook/Customer Shopping Analysis/customer_shopping_behaviour.csv")
```

```
df.head(5)
```

	customer_id	age	gender	item_purchased	category	purchase_amount	location	size	color	season	...	subscription_status	shipping_type	discount_appl
0	2093	42	Male	Sandals	Footwear	45	Minnesota	M	Silver	Fall	...	No	Next Day Air	
1	1117	67	Male	Backpack	Accessories	96	Hawaii	M	Charcoal	Winter	...	No	Express	
2	953	36	Male	Hat	Accessories	71	Idaho	S	Red	Fall	...	Yes	2-Day Shipping	
3	1528	26	Male	Hat	Accessories	75	Wisconsin	L	Red	Fall	...	No	2-Day Shipping	
4	2483	69	Male	Handbag	Accessories	87	Arizona	S	Peach	Winter	...	No	Next Day Air	

- Initial Exploration: Used df.info() to check structure and .describe() for summary statistics.

```
#   Column                   Non-Null Count  Dtype
--  ------                   --------------  -----
0   customer_id              3900 non-null   int64
1   age                      3900 non-null   int64
2   gender                   3900 non-null   object
3   item_purchased           3900 non-null   object
4   category                 3900 non-null   object
5   purchase_amount          3900 non-null   int64
6   location                 3900 non-null   object
7   size                     3900 non-null   object
8   color                    3900 non-null   object
9   season                   3900 non-null   object
10  review_rating            3900 non-null   float64
11  subscription_status      3900 non-null   object
12  shipping_type            3900 non-null   object
13  discount_applied         3900 non-null   object
14  promo_code_used          3900 non-null   object
15  previous_purchases       3900 non-null   int64
16  payment_method           3900 non-null   object
17  frequency_of_purchases   3900 non-null   object
18  date                     3900 non-null   object
19  age_group                3900 non-null   object
20  purchase_frequency_days  3900 non-null   int64
types: float64(1), int64(5), object(15)
emory usage: 640.0+ KB
```

	customer_id	age	purchase_amount	review_rating	previous_purchases	purchase_frequency_days
count	3900.000000	3900.000000	3900.000000	3900.000000	3900.000000	3900.000000
mean	1950.500000	44.068462	59.764359	3.750089	25.351538	89.133077
std	1125.977353	15.207589	23.685392	0.713590	14.447125	119.037566
min	1.000000	18.000000	20.000000	2.500000	1.000000	7.000000
25%	975.750000	31.000000	39.000000	3.100000	13.000000	14.000000
50%	1950.500000	44.000000	60.000000	3.760870	25.000000	30.000000
75%	2925.250000	57.000000	81.000000	4.400000	38.000000	90.000000
max	3900.000000	70.000000	100.000000	5.000000	50.000000	365.000000

- **Missing Data Handling** Missing values in the **review_rating** column were handled using category-wise average ratings.If category averages were unavailable, the overall dataset average was used.
  This ensured no null values while maintaining meaningful customer insights
- **Column Standardization:** Renamed columns to snake case for better readability and documentation.
- **Feature Engineering:**
  - Created age_group column by binning customer ages.
  - Created purchase_frequency_days column from purchase data.

📌 Business Problem Statement

In today's competitive retail environment, understanding customer behavior is essential for improving profitability, retention, and long-term growth. The business currently collects large volumes of customer transaction data, but lacks a structured analytical approach to convert this data into meaningful business insights. The company is facing challenges in identifying:

1. Which product category has consistent buying across all age groups?

2. Which month shows sudden spike in new customers?
3. Which age group has high purchase frequency but low spending amount?
4. Which product categories become less profitable when discounts are applied?
5. What factors (age, gender, or category) most strongly influence customer spending?
6. Which customer group should be targeted for premium memberships based on profitability and loyalty? Additionally, the business does not have clear visibility into:
7. Which age group is the most loyal even if their spending is low?
8. How does purchase frequency differ by age group and product category?
9. How do customer types differ in sales, behavior, and engagement?
10. Which customer group contributes most to subscriptions?

Without this understanding, the company risks inefficient marketing spend, poorly targeted promotions, ineffective discount strategies, and missed opportunities for customer lifetime value growth. Business Problem Statement

Dashboard | Power BI

Key Insights & Findings

- The company is getting fewer new customers.
- Clothes category is being bought more by the young adults.
- The company has got 1053 subscribed customer.
- They are getting more sales from their loyal customers which is $162690 and the average age of the loyal customers is 44.
- Although The company has 2721 loyal customers but only 759 customers have subscribed them.
- Most purchased product by loyal customer is jewelry.
- The company has got new customer in March and August and very less in October.
- Male gender is contributing more in sales than Female which is $1,57,890.
- They have got more Young Adult age-group customer and in that 699 YA customers are loyal but Middle aged age-group has more subscribers.
- Each product is doing great after applying discounts.
- Sandle is Top rated product with 4.0 ratings and Bottom Rated Product is Short with 3.5 ratings

Recommendations

- They Should focus more on promoting and Advertising their Company so that other people get to know about the company and they will get new customers.

- Also they should be giving monthly and seasonal discounts, this will attract more new customers as female customers are contributing less in sales so this will help.
- All the Categories are doing good after discount Applied so they can go with this strategy or they can give other discounts option to their customers.
- 2847 customers have not taken the subscription so they should add more values or offers in their subscription plan.
- loyal customers should be targeted for premium membership specially Young Adult age-group.